25
28
22
32
36
20
40
100 mm
1.0
1.1
1.25
1.4
1.6
2.5
2.8
2.2
3.2
3.6
2.0
4.0
1.8
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm
A5

DOCUMENT RESUME

ED 295 644 IR 013 354

AUTHOR Harvey, Francis A.
TITLE Applications of Artificial Intelligence to Information Search and Retrieval: The Development and Testing of an Intelligent Technical Information System.
PUB DATE 19 Jan 88
NOTE 9p.; In: Proceedings of Selected Research Papers presented at the Annual Meeting of the Association for Educational Communications and Technology (New Orleans, LA, January 14-19, 1988). For the complete proceedings, see IR 013 331.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer System Design; Database Management Systems; *Databases; Engineering; *Expert Systems; *Information Retrieval; *Information Systems; *Optical Data Disks; Programing
IDENTIFIERS Relational Approach (Information Retrieval)

ABSTRACT

This paper describes the evolution and development of an intelligent information system, i.e., a knowledge base for steel structures being undertaken as part of the Technical Information Center for Steel Structures at Lehigh University's Center of Advanced Technology for Large Structural Systems (ATLSS). The initial development of the Technical Information Center is discussed, and the decisions on hardware and software made in the first year are reviewed. In particular, the factors instrumental in the selection of Turbo Prolog as the programming language are detailed. The computer-based bibliographic database designed during the first year of the project and now under development is described, and lessons from the experience and their impact on the future direction of the project are discussed. Projected future activities--including testing of the bibliographic system, expansion of that system into a complete knowledge based system, and conversion of the system to an optical disc (CD-ROM) storage format--and potential research questions associated with developing and using intelligent information systems are also discussed. (EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295644

Applications of Artificial Intelligence to Information Search and Retrieval: The Development and Testing of an Intelligent Technical Information System

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Francis A. Harvey

College of Education
Lehigh University
Bethlehem, Pennsylvania 18015-4793

A paper presented at the annual conference of the

Association for Educational Communications and Technology

New Orleans, Louisiana

19 January, 1988

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael R. Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR013354

Abstract

Problems associated with computer-based information access as currently practiced include the sheer volume of information available, the rapidly changing nature of that information, and the difficulty individuals face in attempting to evaluate all knowledge relevant to a particular problem. The increasing availability of large data bases on optical disc has exacerbated the problem, since search and retrieval technology has not kept pace with advances in data capacity. Intelligent data bases, which attempt to combine techniques used in artificial intelligence for knowledge representation and the development of knowledge-based systems with the highly refined data base techniques already developed to store large amounts of carefully structured and organized data, provide a potential solution to the problem of information overload faced by professionals in all fields.

This paper describes the evolution and development of one information system, the knowledge base for steel structures being developed as part of the Technical Information Center for Steel Structures (TICSS) at Lehigh University's Center for Advanced Technology for Large Structural Systems (ATLSS), which is funded by the National Science Foundation. The paper discusses the initial development of the Technical Information Center and reviews the decisions on hardware and software made in the first year as the project goals were refined, particularly the selection of Turbo Prolog as the programming language. The computer-based bibliographic data base designed during the first year of the project and now under development is described, and lessons gained from the experience and their impact on the future direction of the project are discussed. Projected future activities -- including testing of the bibliographic system, expansion of that system into a complete knowledge based system, and the conversion of the system to optical disc (CD/ROM) storage format -- and potential research questions associated with developing and using "intelligent" information systems are also discussed.

Applications of Artificial Intelligence to Information Search and Retrieval: The Development and Testing of an Intelligent Technical Information System

Introduction: Current Problems

Professionals in nearly every area of education, business, and industry are challenged by a flood of raw information which can overwhelm human comprehension. Computerized data bases now exist which provide access to vast amounts of information, and the amount of information is growing and changing rapidly, approximately doubling every eight years. The introduction of optical disc (CD/ROM) storage technology as a microcomputer peripheral has expanded access to information and at the same time exacerbated the problems of information management and utilization.

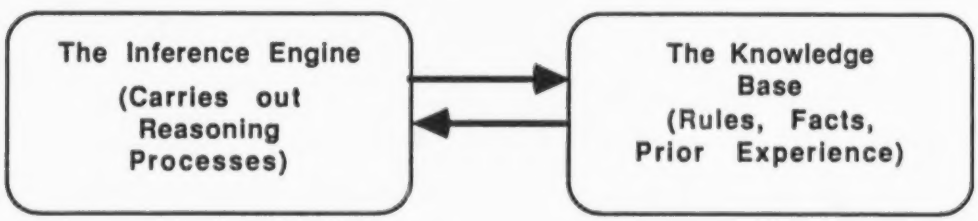
A Potential Solution to Information Overload: Intelligent Data Bases

A potential solution to the problem of information overload exists in the form of the development of "intelligent data bases." Intelligent data bases attempt to combine techniques used in artificial intelligence for knowledge representation and the development of knowledge-based systems with existing data base management techniques developed to store large amounts of data.

Intelligent data bases take advantage of the existing power and flexibility of traditional relational data base management systems (DBMS). Existing DBMS's have the advantage that they are already developed and highly refined, that they can store large amounts of information, and that the information is carefully structured and well organized.

Intelligent data bases apply the technology of expert systems to help researchers identify and focus attention on information relevant to a particular problem. An expert system is "a computing system capable of representing and reasoning about some knowledge-rich domain, such as internal medicine or geology, with a view to solving problems and giving advice" (Jackson, 1986, 1). Figure 1 illustrates the two major components of an expert system, the inference engine and the knowledge base, and the interrelationships between those components.

Figure 1
The Structure of an Expert System



Both the specific information about the domain (that is, facts) and the process by which a human "expert" would solve problems in that domain, are represented symbolically in an expert system. Techniques used to represent knowledge include the use of production rules, structured objects, and predicate logic. For example, in an intelligent data base rules such as constraint checks or checks on the integrity of information referenced by the record can be associated with a data base record, resulting in a data base made up of active or intelligent objects.

A Case Study: A Knowledge Base for Steel Structures

Development of one such intelligent data base is currently in progress at the Technical Information Center for Steel Structures (TICSS) at Lehigh University (Harvey, Stewart, and Beedle, 1987). TICSS is part of the Center for Advanced Technology for Large Structural Systems (ATLSS), an Engineering Research Center at Lehigh funded by the National Science Foundation. The objective of TICSS is to identify the knowledge pertinent to all aspects of steel structures and to make it available for use. The knowledge base will include bibliographical material, technical data, and such other resources as research in progress, case studies, and the particulars about major projects. Professor Lynn S. Beedle, Civil Engineering Department, and the author are principal investigators for the project. TICSS faculty and students come from both the College of Education and the College of Engineering and Applied Science, and work jointly to develop and test applications of technology which will facilitate access by both students and practicing engineers to state-of-the-art information in a variety of forms.

Planning TICSS: Selecting Appropriate Hardware

Planning decisions during the first year of the TICSS project dealt with determining the scope of information to be collected, and identifying the suitable hardware and software environment. MS-DOS microcomputers were selected because most professional engineers used this system and TICSS was conceived as a distributed rather than centralized information system.

Software Selection: Turbo Prolog as the Programming Environment

Turbo Prolog (Borland, 1986) was selected as the programming language in which the initial information subsystems would be developed. Prolog (from PROgramming in LOGic) is a computer language based on a description of human reasoning in terms of formal logic. While most programming languages solve problems using procedures which prescribe how to solve the problem, Prolog uses statements describing objects and the logical relations between them (in the form of facts and rules) to describe what the problem is. The logical statements are interpreted by the computer as programs.

Turbo Prolog was selected as the programming environment to take advantage of its advertised built-in capabilities for creating and manipulating relational data bases. In addition, the information system could be extended by adding more facts and inferential rules to the Prolog program. Development of a knowledge-based system would therefore involve an evolution of the existing system, rather than starting over in a new environment, with the associated problems of converting data from one format to another. This decision has proved to have a major impact on the direction and rate of progress of the TICSS project.

Development of a Computer-Based Bibliographic Data Base

As a first step in developing an intelligent data base, a computer-manipulable information base of bibliographic information and other reference information on steel structures has been designed and is now nearing completion. Materials for the information base were obtained from existing hard copy and machine-readable information bases currently in the library of Lehigh's Fritz Engineering Laboratory, from other sources at Lehigh University, and from the American Institute for Steel Construction (AISC).

As noted above, the bibliographic system was designed in such a way that it could become part of a comprehensive knowledge-based system. The system also was designed for later conversion to optical-disc storage media (CD/ROM, CD-I, or other formats as they become available).

Description of the Bibliographic Data Base

The TICSS bibliographic data base was designed to accomodate information on fourteen different types of documents with a total of 35 fields. The system uses screen templates to guide data entry, and was designed to be very flexible. For example, the bibliographic data base was designed to accomodate documents with any number of authors or editors; it permitted the entry of any number of keywords; and abstracts and comments could be of any length. As noted previously, the original version of the bibliographic system was designed to utilize the built-in relational data base capabilities of Turbo Prolog; thus, the information was stored in sixteen separate files.

A prototype of the relational data base was developed and tested during the spring and summer of 1987, and the program is now being extensively revised based on the experiences with the prototype. The following section will discuss the positive and negative outcomes of experience with the prototype, and describe how plans for the development of future TICSS information systems were modified based on that experience.

Lessons Gained from Experience with the Prototype

Programming in logic rather than in procedures was a very different experience and proved challenging. Programmers had to get used to the different way in which problems had to be conceptualized in designing a Prolog program. Once this cognitive shift was accomplished, however, the programming expertise gained proved valuable in planning and designing the expert system to be developed as the next phase of the TICSS project.

Although Prolog is a very efficient environment for describing and programming facts and logical relationships, some programming tasks which are straightforward in other programming languages turned out to be cumbersome in Prolog. Programming data entry screens, for example, took much longer than anticipated. This situation was improved considerably when the Turbo Prolog Toolbox (Borland, 1987) became available in late Spring, 1987. The Toolbox consists of an extensive set of utilities to facilitate the development of procedures for user input, screen design and layout, menus, graphics, communication with remote devices, and conversion of data from other systems. Extensive use of the Turbo Toolbox procedures is being made in revising the bibliographic data base program.

Prolog is a list-oriented language, so project programmers were able to capitalize on their previous expertise with other list-oriented languages. Designing the data structure as lists and lists of lists, and developing programs to manipulate those lists, was therefore a relatively straightforward task.

A major disappointment was the speed with which the Turbo Prolog program was able to carry out data access and retrieval by means of its built-in relational data base procedures. Turbo Prolog is a compiled language, and executes programs as machine language. However, even with a small number of records (approximately twenty in the original prototype version), writing to files during data entry was very slow, and search and retrieval processes seemed to take forever.

It was clear that, with the approach taken with the prototype software, Turbo Prolog would not be able to meet project needs. Thorough review and analysis of the prototype program and the design specifications of the bibliographic data base were then undertaken. Options considered included changing completely to a commercial relational data base manager such as DBASE III, changing to a more powerful hardware environment which would support the use of more powerful versions of Prolog, and re-designing the program within the Turbo Prolog environment. The latter alternative was judged to best meet project needs, and a totally redesigned and rewritten version of the bibliographic data entry program is now nearing completion. The new version uses one large ASCII archival file instead of multiple relational files for data storage, and makes extensive use of index files and search algorithms developed by project programmers to provide fast and efficient Boolean searching of the archival file. Turbo Toolbox procedures for screen management are being extended to provide a friendlier user interface.

Impact on the Future Direction of the Project

As a result of the experience with the Turbo Prolog prototype, we are exploring the use of a variety of environments for development of additional information subsystems for TICSS. For example, a data base of computer software for engineering applications and a mechanism for identifying current research will be developed in DBASE III+ and procedures for converting DBASE III+ data into a form compatible with the Prolog information system will be developed. A change of both hardware and software environments is actively being considered since the experience of the project has demonstrated that the original environment is probably too limited to meet the needs of the Technical Information Center for Steel Structures.

Future Project Activities

Objectives for the second year of the project are to test the extensively revised bibliographic data base with a variety of users, and to expand the range and usefulness of the data base both by entering additional bibliographic information and by extending the range of types of information which will be included. Eventually the full knowledge base could include information on "documents" of the following types:

- bibliographies (reports, theses, journal articles, books, and design guides)
- technical data
- case studies (and digests)
- current research activities
- slides
- inventory of "large structures"
- project descriptions
- directory of professional specialists
- computer software
- films
- audio- and videotapes
- research results.

At the same time, the incorporation of all data bases into true knowledge-based systems will be explored. Approaches now being considered for combining traditional data base management system capabilities with the power of knowledge-based systems are illustrated in Figure 2 (see following page). These include:

- Adopting or developing a system in which the capabilities of the data base management system (DBMS) are included within a knowledge-based system (KBS);
- A system wherein the KBS accesses the DBMS after a search query has been formulated, that is, where the KBS acts as an "intelligent front end" to the DBMS;
- Utilizing an oversight management system which accesses both the KBS and the DBMS as appropriate and coordinates the exchange of information between the two systems.

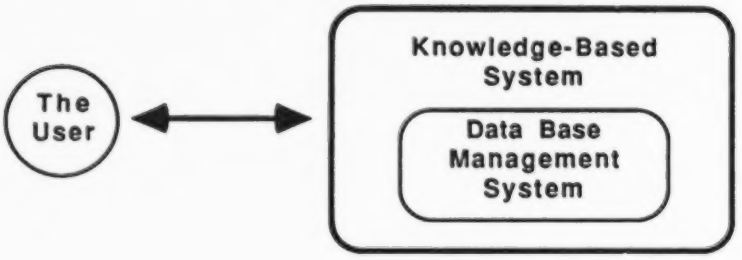
During the second year of the project a plan for converting the system to optical disc (CD/ROM) storage will be developed. In addition, a variety of research projects are now being planned to support these efforts.

Objectives for the third year are to continue to expand the scope and contents of the system and to complete the development of the system for steady-state operation, to complete conversion to optical disc format, and to implement a marketing plan.

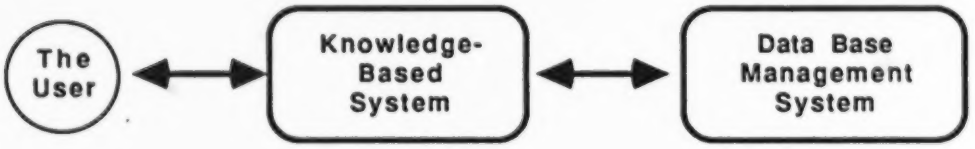
Figure 2

Design Approaches for Intelligent Data Base Systems

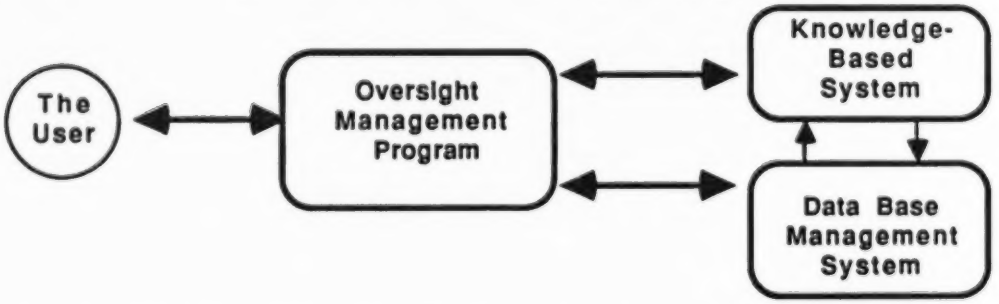
1. Include Data Base Management System (DBMS) Capabilities within Knowledge-Based System (KBS)



2. Knowledge-Based System acts as "Intelligent Front End" to Data Base Management System



3. Use of Oversight Management Program To Control and Coordinate Access to and Between Knowledge-Based System and Data Base Management System



Potential Research Questions Concerning "Intelligent" Information Systems

The advent of new systems of information access opens up a wide range of research questions related to the representation and utilization of knowledge. The following list of potential research questions is representative but by no means exhaustive.

- A comparison of the relative effectiveness of representing knowledge to the user in different ways, for example by the use of icons compared with menus, or graphically compared with tabular data;
- Identification of more effective ways to make use of the expertise of the professional expert who is intimately familiar with the information in the system and who knows

how to apply given criteria to select appropriate subsets of the information system. Translating the insights of such an expert to a set of instructions in a computer program is a major component of the development of a knowledge-based system.

- Investigation of how people develop and utilize higher-level search strategies. For example, strategies which utilize Boolean logic rules (intersection, union, negation) and combinations of Boolean rules (for example, metal fatigue AND welding techniques BUT NOT bridges) to select appropriate information from a large data base.
- Differences in cognitive style and information processing techniques utilized by people when working with computer-based (on-screen) information compared to traditional print materials.

Summary

This paper has described the ongoing development of a project through which information can be more thoroughly comprehended and more properly utilized through the application of artificial intelligence techniques. The project illustrates one way in which artificial intelligence can be applied effectively to problems of the real world.

Instructional developers and researchers in educational technology, as well as all educators, face the problems caused by an explosion of information related to their profession. The advent of new systems of information access opens up a wide range of research questions related to the representation and utilization of knowledge. The potential effectiveness of emerging information systems is enormous. Educational researchers can play a central role in helping these systems meet their potential by providing guidance in terms of thorough and valid research on issues related to the design and utilization of "intelligent" information systems. Artificial intelligence, therefore, offers new opportunities for educational research at the same time if offers potential support in identifying and utilizing professional information. We appear to be at the beginning of a new era, an "Age of Information." The implications of the changes that are rapidly taking place around us have not been fully recognized. The investigation of how professionals access and make use of the vastly increased amount of information available is a fruitful area of study.

References

Borland International (1986). *Turbo Prolog owner's handbook*. Scotts Valley, CA: Author.

Borland International (1987). *Turbo Prolog toolbox user's guide and reference manual*. Scotts Valley, CA: Author.

Gallaire, H., and Minker, J. (1978). *Logic and data bases*. New York: Plenum Press.

Harmon, P., and King, D. (1985). *Expert systems: Artificial intelligence in business*. New York: John Wiley and Sons.

Harvey, F. A., Stewart, G. S., and Beedle, L.S. (1987). *Technical Information Center for Steel Structures (TICSS): System management document* (ATLSS Rep. No. 87-05). Bethlehem, PA: Center for Advanced Technology for Large Structural Systems.

Jackson, P. (1986). *Introduction to expert systems*. Reading, MA: Addison-Wesley.

Monarch, I. and Carbonell, J. (1987). CoalSORT: A knowledge-based interface. *IEEE Expert*, (Spring), 39-53.

Parsaye, K. (1983). Logic programming and relational databases. *Database Engineering*, 8(4) (December), 20-39.

Schur, S. (1988). The intelligent data base. *AI Expert*, 3:1 (January), 27-34.

Sciore, E., and Warren, D. (1988). Integrating data bases and Prolog. *AI Expert*, 3:1 (January), 38-44.

Sleeman, D. and Brown, J.S., Eds. (1982). *Intelligent tutoring systems*. New York: Academic Press.

Winston, P. H. (1984). *Artificial Intelligence*.. 2nd edition. Reading, MA: Addison-Wesley.